

Performance Analysis of ℓ_1 -synthesis with Coherent Frames

Yulong Liu*, Shidong Li† and Tiebin Mi‡

Abstract

Signals with sparse frame representations comprise a much more realistic model of nature than that with orthonormal bases. Studies about the signal recovery associated with such sparsity models have been one of major focuses in compressed sensing. In such settings, one important and widely used signal recovery approach is known as ℓ_1 -synthesis (or Basis Pursuit). We present in this article a more effective performance analysis (than what are available) of this approach in which the dictionary \mathbf{D} may be highly, and even perfectly correlated. Under suitable conditions on the sensing matrix Φ , an error bound of the recovered signal $\hat{\mathbf{f}}$ (by the ℓ_1 -synthesis method) is established. Such an error bound is governed by the decaying property of $\tilde{\mathbf{D}}_o^* \mathbf{f}$, where \mathbf{f} is the true signal and $\tilde{\mathbf{D}}_o$ denotes the optimal dual frame of \mathbf{D} in the sense that $\|\tilde{\mathbf{D}}_o^* \hat{\mathbf{f}}\|_1$ produces the smallest $\|\tilde{\mathbf{D}}^* \tilde{\mathbf{f}}\|_1$ in value among all dual frames $\tilde{\mathbf{D}}$ of \mathbf{D} and all feasible signals $\tilde{\mathbf{f}}$. This new performance analysis departs from the usual description of the combo $\Phi \mathbf{D}$, and places the description on Φ . Examples are demonstrated to show that when the usual analysis fails to explain the working performance of the synthesis approach, the newly established results do.

Keywords: Compressed sensing, coherent frames, ℓ_1 -synthesis, optimal-dual-based ℓ_1 -analysis.

*Yulong Liu is with the Institute of Electronics, Chinese Academy of Sciences, Beijing, 100190, China. Email: {yulong3.liu@gmail.com}.

†Shidong Li is with the Department of Mathematics, San Francisco State University, San Francisco, CA 94132, USA. Email: {shidong@sfsu.edu}.

‡Tiebin Mi is with the School of Information Sciences, Renmin University of China, 100872, China. Email: {mitiebin@gmail.com}.

1 Introduction

Compressed sensing is a new data acquisition theory which allows that sparse or compressible signals of interest can be recovered from a small number of linear, non-adaptive, and usually randomized measurements [8, 9, 13]. By now, compressed sensing has attracted abundant applications in signal and image processing, see e.g., the two special issues [1], [2] and references therein. Formally, one considers the following measurement model:

$$\mathbf{y} = \Phi \mathbf{f} + \mathbf{z}, \quad (1)$$

where Φ is an $m \times n$ sensing matrix with $m \ll n$ (indicating some significant undersampling) and $\mathbf{z} \in \mathbb{R}^m$ is a noise term modeling measurement error. The goal is to reconstruct the unknown signal $\mathbf{f} \in \mathbb{R}^n$ based on available measurements $\mathbf{y} \in \mathbb{R}^m$.

In standard compressed sensing scenarios, it is usually assumed that the signal \mathbf{f} has a sparse (or nearly sparse) representation in an orthonormal basis. However, a large number of applications in signal and image processing point to problems where \mathbf{f} is sparse with respect to an overcomplete dictionary or a frame rather than an orthonormal basis, see, e.g., [24], [11], [4], and references therein. Examples include, e.g., signal modeling in array signal processing (oversampled array steering matrix), reflected radar and sonar signals (Gabor frames), and images with curves (curvelets), etc. The flexibility of frames is the key characteristic that empowers frames to become a natural and concise signal representation tool. Therefore, it is highly desirable to extend the compressed sensing methodology to redundant dictionaries as apposed to orthonormal bases only, see, e.g., [25], [10], [22]. In such sparse frame representation setting, the signal \mathbf{f} is now expressed as $\mathbf{f} = \mathbf{D}\mathbf{x}$ where $\mathbf{D} \in \mathbb{R}^{n \times d}$ ($n < d$) is a matrix of frame¹ vectors (as columns) that are often rather coherent

¹A set of vectors $\{\mathbf{d}_k\}_{k \in I}$ in \mathbb{R}^n is a *frame* of \mathbb{R}^n if there exist constants $0 < A \leq B < \infty$ such that

$$\forall \mathbf{f} \in \mathbb{R}^n, \quad A \|\mathbf{f}\|_2^2 \leq \sum_{k \in I} |\langle \mathbf{f}, \mathbf{d}_k \rangle|^2 \leq B \|\mathbf{f}\|_2^2,$$

where numbers A and B are called lower and upper frame bounds, respectively. More details about frames can be found in e.g., [12], [18].

in applications, and $\mathbf{x} \in \mathbb{R}^d$ is a sparse coefficient vector. The linear measurements of \mathbf{f} then can be written as

$$\mathbf{y} = \Phi \mathbf{D} \mathbf{x} + \mathbf{z}. \quad (2)$$

Since \mathbf{x} is assumed sparse, the standard way of recovering \mathbf{f} from (2) is known as ℓ_1 -synthesis (or synthesis-based method) [11], [16], [10]. From the measurements, one first finds the sparsest possible coefficient \mathbf{x} by solving an ℓ_1 minimization problem

$$\hat{\mathbf{x}} = \underset{\tilde{\mathbf{x}} \in \mathbb{R}^d}{\operatorname{argmin}} \|\tilde{\mathbf{x}}\|_1 \quad s.t. \quad \|\mathbf{y} - \Phi \mathbf{D} \tilde{\mathbf{x}}\|_2 \leq \epsilon, \quad (3)$$

where $\|\mathbf{x}\|_p$ ($p = 1, 2$) denotes the standard ℓ_p -norm of the vector \mathbf{x} and ϵ is an upper bound of the noise². Then the solution to \mathbf{f} is derived via a synthesis operation, i.e., $\hat{\mathbf{f}} = \mathbf{D} \hat{\mathbf{x}}$.

Although empirical studies show that ℓ_1 -synthesis often achieves good recovery results, the theoretical performance of this method is far from satisfactory. The analytical results in [25] essentially require that the frame \mathbf{D} has columns that are extremely uncorrelated such that the compound matrix $\Phi \mathbf{D}$ satisfies the requirements imposed by the traditional compressed sensing assumptions. However, these requirements are often infeasible when \mathbf{D} is highly coherent. For example, consider a simple case in which $\Phi \in \mathbb{R}^{m \times n}$ is a Gaussian matrix with i.i.d. entries, then $\Phi \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{I}_m)$, where \otimes denotes the Kronecker product and \mathbf{I}_m is an identity matrix of the size m . It is now well known that with very high probability Φ satisfies the restricted isometry property (RIP) [6] provided that m is on the order of $s \log(n/s)$ [3], [8]. Let us now examine $\Phi \mathbf{D}$. It is not hard to show that $\Phi \mathbf{D} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}^* \mathbf{D} \otimes \mathbf{I}_m)$, where $(\cdot)^*$ denotes the transpose operation. If \mathbf{D} is unitary, then $\Phi \mathbf{D}$ has the same distribution as Φ , and hence satisfies the RIP. However, if \mathbf{D} is a coherent frame, then $\Phi \mathbf{D}$ may no longer obey the common RIP since the entries of $\Phi \mathbf{D}$ are correlated. Meantime, the mutual incoherence property (MIP) [14] may not apply either, as it is very hard for $\Phi \mathbf{D}$ to satisfy the MIP as well when \mathbf{D} is highly correlated.

²The extension to the Gaussian noise case is straightforward since with large probability, the Gaussian noise belongs to bounded sets, see, e.g., lemma 5.1 in [5].

The perspective of the results in [25] is that some sufficient conditions are put on the compound matrix $\Phi\mathbf{D}$ such that \mathbf{x} can be recovered accurately, which leads to a good estimate of \mathbf{f} . However, if one is only interested in reconstructing the signal \mathbf{f} and may not care about obtaining a good recovery of \mathbf{x} . As pointed out in [10], when the dictionary \mathbf{D} has two identical columns, it seems impossible to recover a unique sparse coefficient vector \mathbf{x} from the measurements, but we may certainly be able to reconstruct the signal \mathbf{f} accurately. In other words, a good recovery of \mathbf{x} may be unnecessary to guarantee an accurate reconstruction of \mathbf{f} .

We observe in abundant examples that the ℓ_1 -synthesis method is also capable of producing fine approximation of \mathbf{f} without recovering accurate coefficient vector \mathbf{x} . Known analysis results such as [25] would then not be able to explain these fine results by the synthesis approach.

In this article, we present a new performance analysis for the ℓ_1 -synthesis approach (3) in which the dictionary \mathbf{D} may be highly - and even perfectly - correlated. To the best knowledge of the authors, our new results are more effective than what are known and available. Our results do not depend on a good recovery of the coefficients. The basic idea is to establish the equivalence between the ℓ_1 -synthesis approach and the optimal-dual-based ℓ_1 -analysis approach recently proposed in [22]. Then the recovery error bound for the latter will naturally lead to that for the former.

This paper is organized as follows. Section 2 introduces the family of analysis-based approaches which includes the standard ℓ_1 -analysis, the general-dual-based ℓ_1 -analysis, and the optimal-dual-based ℓ_1 -analysis. In Section 3, the equivalence between the ℓ_1 -synthesis and the optimal-dual-based ℓ_1 -analysis is established. The new performance analysis (error bound) for the ℓ_1 -synthesis is then naturally followed from that of the optimal-dual based ℓ_1 -analysis approach. Some numerical experiments are presented in Section 4 to demonstrate the effectiveness of the results obtained in Section 3. These examples show that when the usual analysis fails to explain the working performance of the synthesis approach, our newly established results do. Conclusion remarks are given in Section 5.

2 The Family of Analysis-based Approaches Based on General Dual Frames

Alongside the ℓ_1 -synthesis approach, there is a counterpart that takes an analysis point of view, see e.g., [15], [16], [10]. This alternative finds an estimate of \mathbf{f} directly by solving the problem

$$\hat{\mathbf{f}} = \underset{\mathbf{f} \in \mathbb{R}^n}{\operatorname{argmin}} \|\bar{\mathbf{D}}^* \tilde{\mathbf{f}}\|_1 \quad s.t. \quad \|\mathbf{y} - \Phi \tilde{\mathbf{f}}\|_2 \leq \epsilon, \quad (4)$$

where $\bar{\mathbf{D}}$ denotes the canonical dual frame of \mathbf{D} , i.e., $\bar{\mathbf{D}} = (\mathbf{D}\mathbf{D}^*)^{-1}\mathbf{D}$. Note that if \mathbf{D} is a Parseval frame, then we have $\bar{\mathbf{D}} = \mathbf{D}$.

It is well known by now [16] that when \mathbf{D} is a square and invertible dictionary, the ℓ_1 -analysis and ℓ_1 -synthesis approaches are equivalent. However, when \mathbf{D} is an overcomplete frame, the gap between them exists.

A remarkable performance study of the ℓ_1 -analysis approach (4) in the case of Parseval frames ($\bar{\mathbf{D}} = \mathbf{D}$) was given in [10]. It was shown that, under suitable conditions on the sensing matrix Φ , the solution to (4) is very accurate provided that $\mathbf{D}^*\mathbf{f}$ has rapidly decreasing coefficients. In other words, when the frame coefficient vector $\mathbf{D}^*\mathbf{f}$ is reasonably sparse, ℓ_1 -analysis can be the right method to use.

However, that \mathbf{f} is sparse in terms of \mathbf{D} does not imply $\mathbf{D}^*\mathbf{f}$ is necessarily sparse. In fact, as the canonical dual frame expansion in the case of Parseval frames, $\mathbf{D}^*\mathbf{f} = \mathbf{D}^*\mathbf{D}\mathbf{x}$ has the minimum ℓ_2 -norm by the frame property, see, e.g., [12] and is usually fully populated which is also pointed out in [25]. In other words, the canonical dual frame of \mathbf{D} may be ineffective in sparsifying \mathbf{f} since ℓ_2 -norm tends to spread the coefficients into a large number of small coefficients.

To overcome this difficulty, the standard ℓ_1 -analysis approach (4) has recently been extended to a more general case in which the analysis operator can be any dual frame³ of \mathbf{D} [22]. This leads to

³A frame $\{\tilde{\mathbf{d}}_k\}_{k \in I}$ is an alternative dual frame of $\{\mathbf{d}_k\}_{k \in I}$ if

$$\forall \mathbf{f} \in \mathbb{R}^n, \quad \mathbf{f} = \sum_{k \in I} \langle \mathbf{f}, \tilde{\mathbf{d}}_k \rangle \mathbf{d}_k = \sum_{k \in I} \langle \mathbf{f}, \mathbf{d}_k \rangle \tilde{\mathbf{d}}_k.$$

the following general-dual-based ℓ_1 -analysis approach

$$\hat{\mathbf{f}} = \underset{\mathbf{f} \in \mathbb{R}^n}{\operatorname{argmin}} \|\tilde{\mathbf{D}}^* \mathbf{f}\|_1 \quad s.t. \quad \|\mathbf{y} - \Phi \mathbf{f}\|_2 \leq \epsilon, \quad (5)$$

where columns of $\tilde{\mathbf{D}}$ form a general (and any) dual frame of \mathbf{D} . The performance analysis of the general-dual-based ℓ_1 -analysis approach was also given in [22]. In order to introduce the results, we require the concept of **D**-RIP [10]: An $m \times n$ sensing matrix Φ is said to satisfy the restricted isometry property adapted to \mathbf{D} (abbreviated **D**-RIP) with constant $\delta_s \in (0, 1)$ if

$$(1 - \delta_s) \|\mathbf{v}\|_2^2 \leq \|\Phi \mathbf{v}\|_2^2 \leq (1 + \delta_s) \|\mathbf{v}\|_2^2 \quad (6)$$

holds for all $\mathbf{v} \in \Sigma_s$, where Σ_s is the union of all subspaces spanned by all subsets of s columns of \mathbf{D} . The validity of the **D**-RIP was discussed in e.g., [10], [20]. It was shown in [10] that any $m \times n$ matrix Φ obeying for any fixed $\boldsymbol{\nu} \in \mathbb{R}^n$

$$\Pr \left(\left| \|\Phi \boldsymbol{\nu}\|_2^2 - \|\boldsymbol{\nu}\|_2^2 \right| \geq \delta \|\boldsymbol{\nu}\|_2^2 \right) \leq c e^{-\gamma \delta^2 m}, \quad \delta \in (0, 1) \quad (7)$$

(γ, c are positive constants) will satisfy the **D**-RIP with overwhelming probability provided that m is on the order of $s \log(d/s)$. Many types of random matrices satisfy (7), some examples include matrices with Gaussian, subgaussian, or Bernoulli entries. It has also been shown in [20] that randomizing the column signs of any matrix that satisfies the standard RIP results in a matrix which satisfies the Johnson-Lindenstrauss lemma [19]. Such a matrix would then satisfy the **D**-RIP via (7). Consequently, partial Fourier matrix (or partial circulant matrix) with randomized column signs will satisfy the **D**-RIP since these matrices are known to satisfy the RIP.

With these preliminaries, we now restate the results in [22] as follows.

Theorem 1. [22] *Let \mathbf{D} be a general frame of \mathbb{R}^n with frame bounds $0 < A \leq B < \infty$. Let $\tilde{\mathbf{D}}$ be an alternative dual frame of \mathbf{D} with frame bounds $0 < \tilde{A} \leq \tilde{B} < \infty$, and let $\rho = s/b$. Suppose*

$$\left(1 - \sqrt{\rho B \tilde{B}} \right)^2 \cdot \delta_{s+a} + \rho B \tilde{B} \cdot \delta_b < 1 - 2\sqrt{\rho B \tilde{B}} \quad (8)$$

holds for some positive integers a and b satisfying $0 < b - a \leq 3a$. Then the solution $\hat{\mathbf{f}}$ to (5) satisfies

$$\|\hat{\mathbf{f}} - \mathbf{f}\|_2 \leq C_0 \cdot \epsilon + C_1 \cdot \frac{\|\tilde{\mathbf{D}}^* \mathbf{f} - (\tilde{\mathbf{D}}^* \mathbf{f})_s\|_1}{\sqrt{s}}, \quad (9)$$

where C_0 and C_1 are some constants and $(\tilde{\mathbf{D}}^* \mathbf{f})_s$ denotes the vector consisting the largest s entries of $\tilde{\mathbf{D}}^* \mathbf{f}$ in magnitude (and setting the other to zero).

Theorem 1 shows that if Φ satisfies some proper conditions, e.g., (8), then the solution to (5) is very accurate provided that $\tilde{\mathbf{D}}^* \mathbf{f}$ has rapidly decreasing coefficients. By the definition of the \mathbf{D} -RIP, the condition (8) is independent of the coherence of the dictionary \mathbf{D} . For differently chosen a and b , (8) will give rise to different conditions on the \mathbf{D} -RIP constants δ_{s+a} and δ_b . For instance, if \mathbf{D} is a Parseval frame and $\tilde{\mathbf{D}}$ is its canonical dual frame, i.e., $B\tilde{B} = 1$, then (8) is satisfied whenever $\delta_{2s} < 0.1398$ [22].

With the error bound (9), we can easily see the potential superiority of using alternative dual frames as analysis operators. For clarity, we consider a simple case in which the noise is free, i.e., $\epsilon = 0$. Then the error bound (9) reduces to

$$\|\hat{\mathbf{f}} - \mathbf{f}\|_2 \leq C_1 \cdot \frac{\|\tilde{\mathbf{D}}^* \mathbf{f} - (\tilde{\mathbf{D}}^* \mathbf{f})_s\|_1}{\sqrt{s}}. \quad (10)$$

Clearly, the quality of the bound $\|\tilde{\mathbf{D}}^* \mathbf{f} - (\tilde{\mathbf{D}}^* \mathbf{f})_s\|_1 / \sqrt{s}$ in (10) is measured in terms of how effective $\tilde{\mathbf{D}}^* \mathbf{f}$ is in sparsifying the signal \mathbf{f} with respect to the dictionary \mathbf{D} . To explain, suppose that \mathbf{f} has a sparse representation in \mathbf{D} , i.e., $\mathbf{f} = \mathbf{D}\mathbf{x}$, where \mathbf{x} is a sparse coefficient vector. As discussed before, the canonical dual frame expansion of \mathbf{f} has the minimum ℓ_2 -norm, i.e., $\|\tilde{\mathbf{D}}^* \mathbf{f}\|_2 = \min_{\tilde{\mathbf{x}}: \mathbf{D}\tilde{\mathbf{x}} = \mathbf{f}} \|\tilde{\mathbf{x}}\|_2$, and is ineffective in promoting sparsity in general. On the other hand, when the analysis operator can be any dual frame of \mathbf{D} , it is not hard to imagine that there should be some dual frame of \mathbf{D} , denoted by $\tilde{\mathbf{D}}_S$, such that $\tilde{\mathbf{D}}_S^* \mathbf{f} = \mathbf{x}$. This is due to the fact that all coefficients of a frame expansion of \mathbf{f} in \mathbf{D} should correspond to some dual frame of \mathbf{D} , which really is the spirit of frame expansions. Generally, $\tilde{\mathbf{D}}_S$ is much more effective in sparsifying the signal \mathbf{f} than the canonical dual frame does. Therefore, one may expect a better recovery performance by taking some “proper” alternative dual frame as the analysis operator.

The important question then is how to choose some appropriate dual frame such that the corresponding analysis coefficients are as sparse as possible. Since the true \mathbf{f} is never known before hand in practice, it seems to impossible to explicitly construct some proper dual frame $\tilde{\mathbf{D}}$ such that $\tilde{\mathbf{D}}^*\mathbf{f}$ is sparse without additional priori knowledge about the signal \mathbf{f} . One approach proposed in [22] is by the method of optimal-dual-based ℓ_1 -analysis:

$$\hat{\mathbf{f}} = \underset{\substack{\mathbf{D}\tilde{\mathbf{D}}^*=\mathbf{I}, \mathbf{f}\in\mathbb{R}^n}}{\operatorname{argmin}} \|\tilde{\mathbf{D}}^*\tilde{\mathbf{f}}\|_1 \quad s.t. \quad \|\mathbf{y} - \Phi\tilde{\mathbf{f}}\|_2 \leq \epsilon, \quad (11)$$

where the optimization is performed simultaneously over both all dual frames $\tilde{\mathbf{D}}$ of \mathbf{D} and the feasible signal set. This seemingly complicated optimization problem can be reformulated into a simplified form. Note that the class of all dual frames for \mathbf{D} is given by [21]

$$\tilde{\mathbf{D}} = (\mathbf{D}\mathbf{D}^*)^{-1}\mathbf{D} + \mathbf{W}^*(\mathbf{I}_d - \mathbf{D}^*(\mathbf{D}\mathbf{D}^*)^{-1}\mathbf{D}) = \bar{\mathbf{D}} + \mathbf{W}^*\mathbf{P}, \quad (12)$$

where $\mathbf{P} \equiv \mathbf{I}_d - \mathbf{D}^*(\mathbf{D}\mathbf{D}^*)^{-1}\mathbf{D}$ denotes the orthogonal projection onto the null space of \mathbf{D} and $\mathbf{W} \in \mathbb{R}^{d \times n}$ is an arbitrary matrix. Plug (12) into (11), we obtain

$$(\hat{\mathbf{f}}, \hat{\mathbf{g}}) = \underset{\substack{\mathbf{f}\in\mathbb{R}^n, \mathbf{g}\in\mathbb{R}^d}}{\operatorname{argmin}} \|\bar{\mathbf{D}}^*\tilde{\mathbf{f}} + \mathbf{P}\mathbf{g}\|_1 \quad s.t. \quad \|\mathbf{y} - \Phi\tilde{\mathbf{f}}\|_2 \leq \epsilon, \quad (13)$$

where we have used the fact that when $\tilde{\mathbf{f}} \neq \mathbf{0}$, $\mathbf{g} \equiv \mathbf{W}\tilde{\mathbf{f}}$ can be any vector in \mathbb{R}^d due to the fact that \mathbf{W} is free. Note that if $\mathbf{P}\mathbf{g} \equiv \mathbf{0}$, then (13) reduces to the standard ℓ_1 -analysis approach (4). In [22], an iterative algorithm based on the split Bregman iteration [17] was developed to solve the optimization problem (13) efficiently.

Clearly, the solution to (11) definitely corresponds to that of (5) with some “optimal” dual frame, say $\tilde{\mathbf{D}}_o$ as the analysis operator. The optimality here is in the sense that $\|\tilde{\mathbf{D}}_o^*\hat{\mathbf{f}}\|_1$ achieves the smallest $\|\tilde{\mathbf{D}}^*\tilde{\mathbf{f}}\|_1$ in value among all dual frames $\tilde{\mathbf{D}}$ of \mathbf{D} and feasible signals $\tilde{\mathbf{f}}$ satisfied the constraint in (11). Once $\hat{\mathbf{f}}$ and $\hat{\mathbf{g}}$ are obtained (through solving (13)), it follows from (12) that the analysis operator $\tilde{\mathbf{D}}_o^*$ is given by

$$\tilde{\mathbf{D}}_o^* = \bar{\mathbf{D}}^* + \mathbf{P}\mathbf{W}_o, \quad (14)$$

with \mathbf{W}_o satisfying

$$\hat{\mathbf{g}} = \mathbf{W}_o \hat{\mathbf{f}}. \quad (15)$$

Evidently, the optimal dual frame $\tilde{\mathbf{D}}_o$ depends on the solutions of (13). By utilizing the fact that $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^* \otimes \mathbf{A})\text{vec}(\mathbf{B})$, the above equation (15) is equivalent to

$$(\hat{\mathbf{f}}^* \otimes \mathbf{I}_d) \cdot \text{vec}(\mathbf{W}_o) = \hat{\mathbf{g}}, \quad (16)$$

where $\text{vec}(\mathbf{W}_o)$ denotes the vectorization of the matrix \mathbf{W}_o by stacking the columns of \mathbf{W}_o into a single vector. Evidently, the solution to (16) is non-unique in general since this equation is highly underdetermined with n equations but nd unknowns. The class of solutions to (16) is given by

$$\begin{aligned} \text{vec}(\mathbf{W}_o) &= (\hat{\mathbf{f}}^* \otimes \mathbf{I}_d)^\dagger \hat{\mathbf{g}} + \left(\mathbf{I}_{nd} - (\hat{\mathbf{f}}^* \otimes \mathbf{I}_d)^\dagger (\hat{\mathbf{f}}^* \otimes \mathbf{I}_d) \right) \mathbf{w} \\ &= (\hat{\mathbf{f}} \otimes \mathbf{I}_d) \hat{\mathbf{g}} / \|\hat{\mathbf{f}}\|_2^2 + \left(\mathbf{I}_{nd} - (\hat{\mathbf{f}} \hat{\mathbf{f}}^* \otimes \mathbf{I}_d) \right) \mathbf{w} / \|\hat{\mathbf{f}}\|_2^2, \end{aligned} \quad (17)$$

where \mathbf{A}^\dagger denotes the pseudo-inverse of \mathbf{A} and $\mathbf{w} \in \mathbb{R}^{nd \times 1}$ is an arbitrary vector. In deriving (17), we have used the two facts that $(\mathbf{A} \otimes \mathbf{B})^\dagger = \mathbf{A}^\dagger \otimes \mathbf{B}^\dagger$ and $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, see e.g., [23]. Let $\mathbf{w} = \mathbf{0}$, then (17) reduces to the least square solution of \mathbf{W}_o

$$\mathbf{W}_o^{\text{ls}} = (\hat{\mathbf{f}}^* \otimes \hat{\mathbf{g}}) / \|\hat{\mathbf{f}}\|_2^2. \quad (18)$$

If we choose $\mathbf{W}_o = \mathbf{W}_o^{\text{ls}}$, then (14) becomes

$$\tilde{\mathbf{D}}_o^* = \bar{\mathbf{D}}^* + \mathbf{P} \mathbf{W}_o^{\text{ls}} = \bar{\mathbf{D}}^* + (\hat{\mathbf{f}}^* \otimes \mathbf{P} \hat{\mathbf{g}}) / \|\hat{\mathbf{f}}\|_2^2. \quad (19)$$

It is this form (19) which will be used to construct the optimal dual frame in the numerical experiments.

Figure 1 provides a schematic overview of the family of dual-based ℓ_1 -analysis approaches. For the standard ℓ_1 -analysis approach (4) which uses the canonical dual frame of \mathbf{D} as the analysis operator, the recovered signal $\hat{\mathbf{f}}$ has the smallest $\|\bar{\mathbf{D}}^* \hat{\mathbf{f}}\|_1$ in value among the feasible signal set. While for the optimal-dual-based ℓ_1 -analysis approach (11), the optimization is not only over the

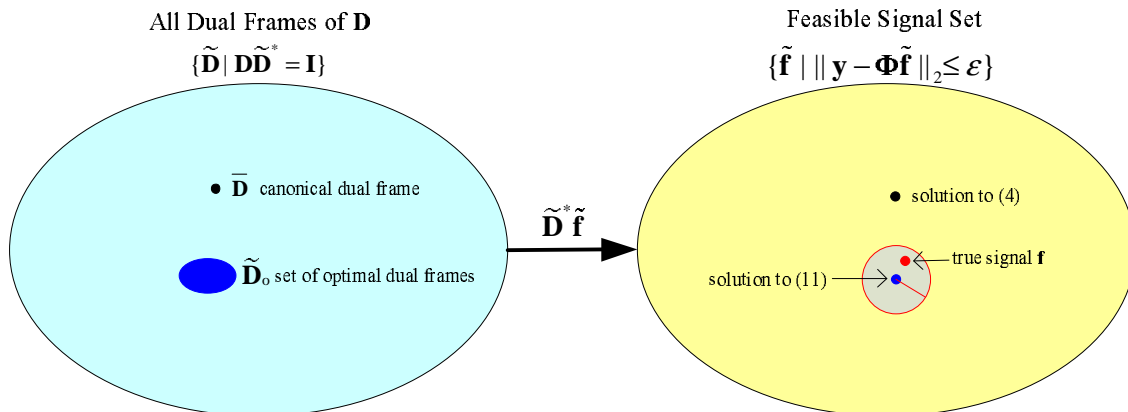


Figure 1: A schematic overview of the family of dual-based ℓ_1 -analysis approaches

feasible signal set but also over all dual frames $\tilde{\mathbf{D}}$ of \mathbf{D} . The recovered signal $\hat{\mathbf{f}}$ and optimal dual frame $\tilde{\mathbf{D}}_o$ (non-unique) produce the smallest $\|\tilde{\mathbf{D}}^* \tilde{\mathbf{f}}\|_1$ in value. When the signal of interest has a sparse representation in a redundant frame, one may expect that the optimal dual frame may be much effective in sparsifying the true signal than the canonical dual frame does. Then a better recovery performance may be achieved by the optimal-dual-based ℓ_1 -analysis approach. Indeed, we have seen that the signal recovery via (11) is much more effective than that of the standard ℓ_1 -analysis approach (4) which uses the canonical dual frame as the analysis operator. Moreover, the optimal-dual-based analysis method provides a new and more effective performance analysis to the ℓ_1 -synthesis approach.

3 Performance Analysis of ℓ_1 -Synthesis

In this section, we present a new performance analysis of the ℓ_1 -synthesis approach. We begin by showing that the ℓ_1 -synthesis and the optimal-dual-based ℓ_1 -analysis approaches are equivalent.

Theorem 2. *ℓ_1 -synthesis and optimal-dual-based ℓ_1 -analysis are equivalent.*

Proof. We start with the optimal-dual-based ℓ_1 -analysis approach as posed in (13). Let $\tilde{\mathbf{x}} = \tilde{\mathbf{D}}^* \tilde{\mathbf{f}} + \mathbf{P}\mathbf{g}$, then we have $\mathbf{D}\tilde{\mathbf{x}} = \tilde{\mathbf{f}}$. Since both $\tilde{\mathbf{f}}$ and \mathbf{g} are free, then $\tilde{\mathbf{x}} \in \mathbb{R}^d$. Put the two facts into (13), we

obtain the ℓ_1 -synthesis method (3). On the other hand, we start from the ℓ_1 -synthesis formulation.

For any $\tilde{\mathbf{x}} \in \mathbb{R}^d$, the following decomposition always holds

$$\begin{aligned}\tilde{\mathbf{x}} &= \tilde{\mathbf{x}}_R + \tilde{\mathbf{x}}_N = \mathbf{D}^*(\mathbf{D}\mathbf{D}^*)^{-1}\mathbf{D}\tilde{\mathbf{x}} + \mathbf{P}\tilde{\mathbf{x}} \\ &= \tilde{\mathbf{D}}^*\mathbf{D}\tilde{\mathbf{x}} + \mathbf{P}\tilde{\mathbf{x}},\end{aligned}$$

where $\tilde{\mathbf{x}}_R$ and $\tilde{\mathbf{x}}_N$ are the components of $\tilde{\mathbf{x}}$ belonging to the row space and the null space of \mathbf{D} , respectively. Define $\tilde{\mathbf{f}} = \mathbf{D}\tilde{\mathbf{x}} \in \mathbb{R}^n$ and $\mathbf{g} = \tilde{\mathbf{x}} \in \mathbb{R}^d$, we can arrive at the optimal-dual-based ℓ_1 -analysis approach and the two methods are equivalent. \square

Remark 1: By taking a geometrical description, it was shown in [16] that any ℓ_1 -analysis problem (with full-rank analysis operator) may be reformulated as an equivalent ℓ_1 -synthesis one. Our results indicate that the reverse is also true. For a given ℓ_1 -synthesis problem, there exist some appropriate analysis operators (e.g., optimal dual frames of \mathbf{D}) such that the corresponding ℓ_1 -analysis problem is equivalent to the ℓ_1 -synthesis one.

With this equivalence, we now establish the error bound of the ℓ_1 -synthesis approach. Since $\tilde{\mathbf{D}}_o$ is some alternative dual frame of \mathbf{D} , i.e., $\mathbf{D}\tilde{\mathbf{D}}_o^* = \mathbf{I}$, a direct application of Theorem 1 leads to the following results.

Theorem 3. *Let \mathbf{D} be a general frame of \mathbb{R}^n with frame bounds $0 < A \leq B < \infty$. Let $\tilde{\mathbf{D}}_o$ be some optimal dual frame of \mathbf{D} defined in (14) with frame bounds $0 < \tilde{A}_o \leq \tilde{B}_o < \infty$, and let $\rho = s/b$. Suppose*

$$\left(1 - \sqrt{\rho B \tilde{B}_o}\right)^2 \cdot \delta_{s+a} + \rho B \tilde{B}_o \cdot \delta_b < 1 - 2\sqrt{\rho B \tilde{B}_o} \quad (20)$$

holds for some positive integers a and b satisfying $0 < b - a \leq 3a$. Then the solution $\hat{\mathbf{f}}$ to (3) (or to (13)) satisfies

$$\|\hat{\mathbf{f}} - \mathbf{f}\|_2 \leq C_0 \cdot \epsilon + C_1 \cdot \frac{\|\tilde{\mathbf{D}}_o^* \mathbf{f} - (\tilde{\mathbf{D}}_o^* \mathbf{f})_s\|_1}{\sqrt{s}}, \quad (21)$$

where C_0 and C_1 are some constants and $(\tilde{\mathbf{D}}_o^ \mathbf{f})_s$ denotes the vector consisting the largest s entries of $\tilde{\mathbf{D}}_o^* \mathbf{f}$ in magnitude.*

Theorem 3 shows that, under suitable conditions on the sensing matrix, the recovered signal $\hat{\mathbf{f}}$ by ℓ_1 -synthesis is very accurate provided that $\tilde{\mathbf{D}}_o^* \mathbf{f}$ has rapidly decreasing coefficients. By the optimality of $\tilde{\mathbf{D}}_o$, one may expect that $\tilde{\mathbf{D}}_o$ will promote high sparsity in the frame expansion of the signal \mathbf{f} . Indeed, as we shall see in the numerical experiments, $\tilde{\mathbf{D}}_o$ is much more effective in sparsifying \mathbf{f} than the canonical dual frame does. Consequently, in comparison to the standard ℓ_1 -analysis approach, a better signal recovery is often achieved by ℓ_1 -synthesis.

More importantly, this new performance analysis result is capable of explaining examples of successful solutions and fine approximations by the ℓ_1 -synthesis approach while the recovered coefficient vector \mathbf{x} is nowhere near its true value. Known performance analysis results would not have such capacity.

4 Numerical Results

In this section, we present some numerical experiments to demonstrate the effectiveness of the performance analysis results for ℓ_1 -synthesis. In these experiments, we use two types of frames: Gabor frames and a concatenation of the coordinate and Fourier bases. The sensing matrix Φ is a Gaussian matrix with $m = 32$, $n = 128$. Since the dependence on the noise in the error bound (21) is optimal and for the purpose of clarity, we only consider the noise-free case. Both ℓ_1 -analysis and ℓ_1 -synthesis problems are solved by the algorithm developed in [22] because the returned auxiliary variable $(\mathbf{P}\mathbf{g})$ by this algorithm can be used to construct the optimal dual frame $\tilde{\mathbf{D}}_o$ (19). For completeness of this paper, this algorithm is included in Appendix A. We set $\lambda = \mu = 1$, $tol = 10^{-12}$, $nInner = 5$, and $nOuter = 100$ in this algorithm for all experiments.

Example 1: Gabor Frames. Recall that for a window function g and positive time-frequency shift parameters α and β , the Gabor frame is given by

$$\{g_{l,k}(t) = g(t - k\alpha)e^{2\pi i l \beta t}\}_{l,k}. \quad (22)$$

For many practical applications such as radar and sonar, the received signal f often has the form

$$f(t) = \sum_{k=1}^s a_k g(t - t_k) e^{i\omega_k t}. \quad (23)$$

Evidently, f is sparse with respect to a Gabor frame. In this experiment, we construct a Gabor dictionary with Gaussian windows, oversampled by a factor of 30 so that $d = 30 \times n = 3840$. The tested signal \mathbf{f} is sparse with respect to the constructed Gabor frame with sparsity $s = \text{ceil}(0.2 \times m) = 7$. The positions of the nonzero entries of the coefficient vector \mathbf{x} are selected uniformly at random, and each nonzero value is sampled from standard Gaussian distribution.

Figure 2 (a) shows that when \mathbf{D} is highly coherent with coherence⁴ $\mu(\mathbf{D}) = 0.9934$, the recovered coefficients by the ℓ_1 -synthesis are disappointing (with a relative error $\|\bar{\mathbf{x}} - \mathbf{x}\|_2 / \|\mathbf{x}\|_2 = 0.9039$). However, the signal recovered by the ℓ_1 -synthesis is nevertheless quite acceptable (with a relative error equal to 0.0845), see Figure 2 (b). This example tells us that a good recovery of the coefficients \mathbf{x} may be unnecessary to guarantee a fine reconstruction of the signal \mathbf{f} . This phenomenon is explainable by the new performance analysis result, but not by performance results based on the accuracy of the recovery of the coefficient vector \mathbf{x} .

Figure 2 (b) also shows that the signal recovery via ℓ_1 -synthesis is much better than that of ℓ_1 -analysis (relative error: 0.0845 vs. 0.3445). This is because the optimal dual frame $\tilde{\mathbf{D}}_o$ is much more effective in promoting sparsity in the frame expansion of \mathbf{f} than the canonical dual frame $\bar{\mathbf{D}}$ does. Figure 2 (c) compares the largest 100 coefficients (in magnitude) of $\bar{\mathbf{D}}^* \mathbf{f}$ and $\tilde{\mathbf{D}}_o^* \mathbf{f}$, where $\tilde{\mathbf{D}}_o^*$ is determined by (19).

Example 2: Concatenations. When signals of interest are sparse over several orthonormal bases (or frames), it is natural to use a dictionary \mathbf{D} consisting of a concatenation of these bases (or frames). In this experiment, we consider a dictionary consisting of the coordinate and Fourier bases, i.e., $\mathbf{D} = [\mathbf{I}, \mathbf{F}] / \sqrt{2}$. The tested signal \mathbf{f} is a linear combination of spikes and sinusoids, i.e.,

⁴The coherence of the dictionary \mathbf{D} is defined as $\mu(\mathbf{D}) = \max_{j \neq k} \frac{|\langle \mathbf{d}_j, \mathbf{d}_k \rangle|}{\|\mathbf{d}_j\|_2 \|\mathbf{d}_k\|_2}$, where \mathbf{d}_j and \mathbf{d}_k denote columns of \mathbf{D} .

We say that \mathbf{D} is incoherent if $\mu(\mathbf{D})$ is small.

$\mathbf{f} = \mathbf{f}_1 + \mathbf{f}_2 = \mathbf{x}_1 + \mathbf{F}\mathbf{x}_2$. The sparsity of both \mathbf{x}_1 and \mathbf{x}_2 is equal to 4. Again, the positions of the nonzero entries of both \mathbf{x}_1 and \mathbf{x}_2 are selected uniformly at random, and each nonzero value is sampled from standard Gaussian distribution.

Figure 3 (a) and (b) show that when \mathbf{D} is incoherent (with coherence $\mu(\mathbf{D}) = 1/\sqrt{n} = 0.0884$), the ℓ_1 -synthesis approach not only recovers the signal \mathbf{f} but also the coefficient vector \mathbf{x} accurately.

Figure 3 (b) also shows that ℓ_1 -analysis fails in recovering the signal with a relative error at 0.8143. Such a failure is not surprising since $\bar{\mathbf{D}} = \mathbf{D}$ is ineffective in sparsifying the true signal \mathbf{f} , see Figure 3 (c). By contrast, $\tilde{\mathbf{D}}_o^* \mathbf{f}$ decays very quickly, which guarantees the good recovery for the signal by ℓ_1 -synthesis.

5 Conclusions

This paper has presented a novel performance analysis for ℓ_1 -synthesis in which the dictionary may be highly coherent. Our approach was to show the equivalence between ℓ_1 -synthesis and optimal-dual-based ℓ_1 -analysis. With this equivalence, the signal recovery error bound for both could be established by using the results in [22]. Finally, the results obtained in this paper were validated via numerical experiments.

A Split Bregman Iteration for optimal-dual-based ℓ_1 -analysis

This appendix includes the split Bregman iteration for optimal-dual-based ℓ_1 -analysis in which

- \mathbf{f} : the recovered signal;
- \mathbf{x} : the recovered coefficient vector;
- $\mathbf{P}\mathbf{g}$: the auxiliary variable used to construct the optimal dual frame of \mathbf{D} ;
- $\text{shrink}(\cdot)$: denotes the element-wise soft shrinkage operation;

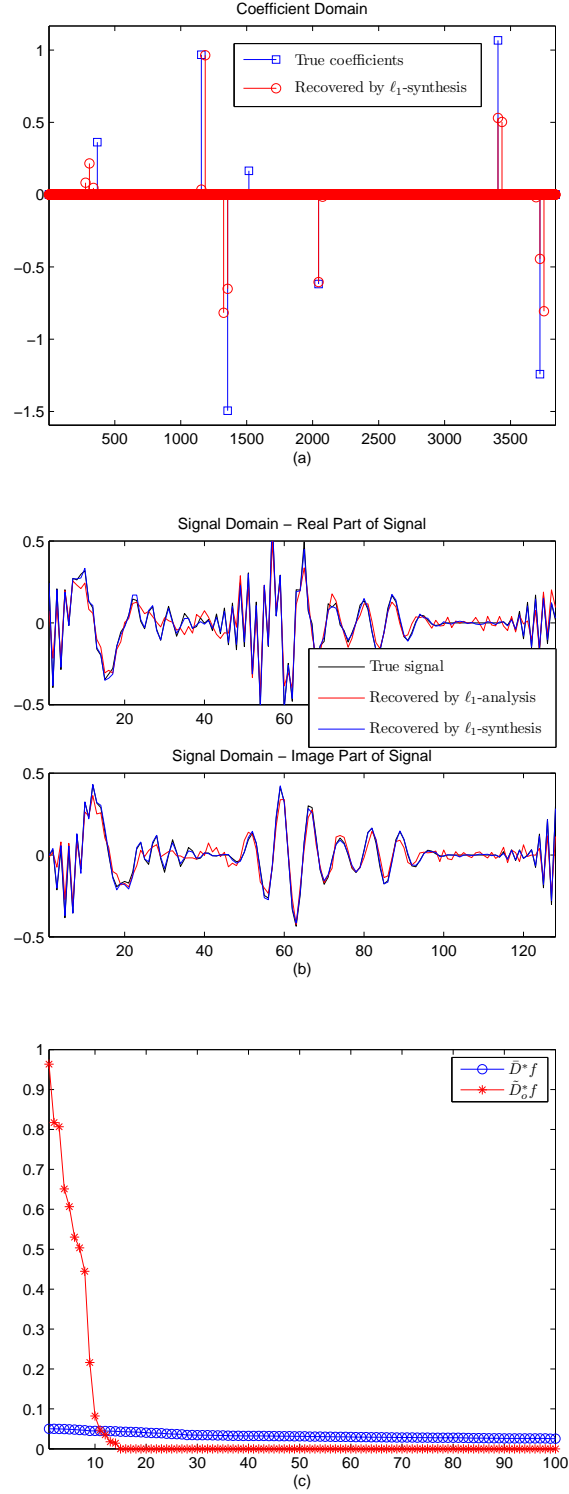


Figure 2: \mathbf{D} = Gabor frame. (a): recovery in coefficient domain by ℓ_1 -synthesis (relative error: 0.9039) with the relative error defined as $\|\hat{\mathbf{x}} - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$. (b): recovery in signal domain by ℓ_1 -analysis (relative error: 0.3445) and ℓ_1 -synthesis (relative error: 0.0845) with the relative error defined as $\|\hat{\mathbf{f}} - \mathbf{f}\|_2 / \|\mathbf{f}\|_2$. (c): The largest 100 coefficients of the coefficient vector $\bar{\mathbf{D}}^* \mathbf{f}$ and $\tilde{\mathbf{D}}_o^* \mathbf{f}$ in magnitude.

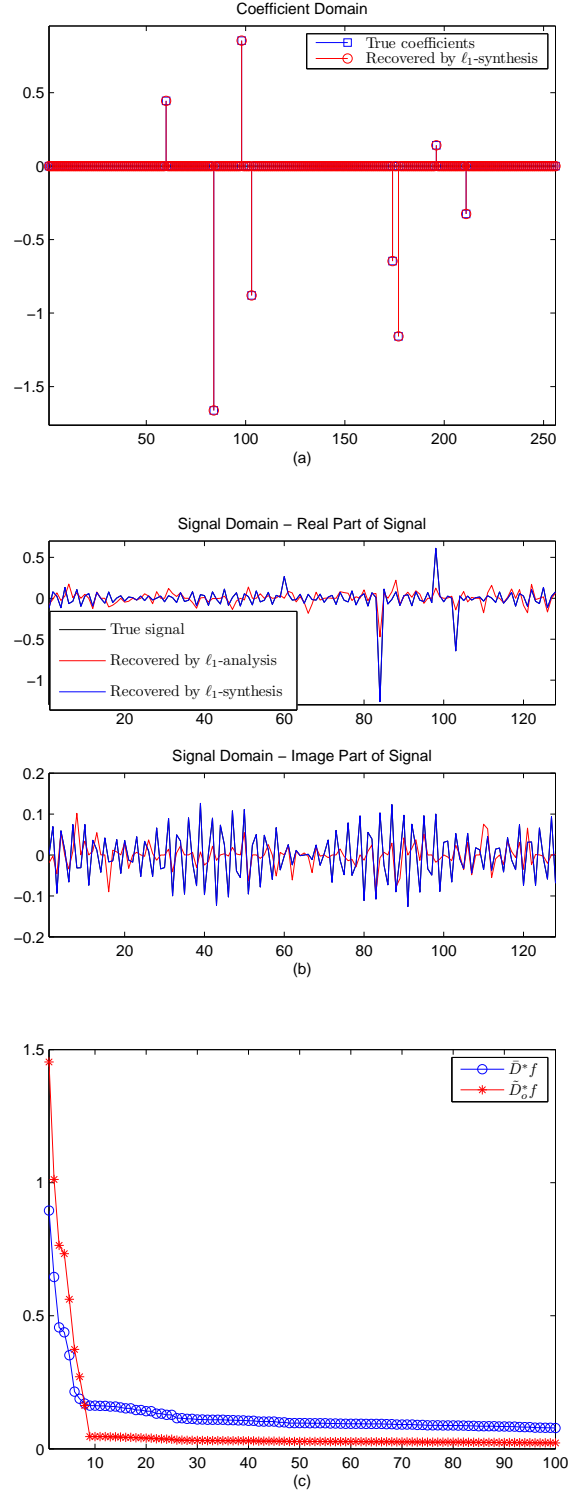


Figure 3: $\mathbf{D} = [\mathbf{I}, \mathbf{F}]/\sqrt{2}$. (a): recovery in coefficient domain by ℓ_1 -synthesis (relative error: 7.7681×10^{-6}) with the relative error defined as $\|\hat{\mathbf{x}} - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$. (b): recovery in signal domain by ℓ_1 -analysis (relative error: 0.8143) and ℓ_1 -synthesis (relative error: 6.7911×10^{-6}) with the relative error defined as $\|\hat{\mathbf{f}} - \mathbf{f}\|_2 / \|\mathbf{f}\|_2$. (c): The largest 100 coefficients of the coefficient vector $\bar{\mathbf{D}}^* \mathbf{f}$ and $\tilde{\mathbf{D}}_o^* \mathbf{f}$ in magnitude.

- $(\cdot)^{new}$: denotes either $(\cdot)^{k+1}$ if it is available or $(\cdot)^k$ otherwise.

Algorithm 1: Split Bregman Iteration for optimal-dual-based ℓ_1 -analysis

Initialization: $\mathbf{f}^0 = \mathbf{0}$, $\mathbf{x}^0 = \mathbf{b}^0 = \mathbf{P}\mathbf{g}^0 = \mathbf{0}$, $\mathbf{c}^0 = \mathbf{0}$, $\mu > 0, \lambda > 0, nOuter, nInner, tol$;

Output: \mathbf{f} , \mathbf{x} , $\mathbf{P}\mathbf{g}$;

while $k < nOuter$ **and** $\|\Phi\mathbf{f}^k - \mathbf{y}\|_2 > tol$ **do**

for $n = 1 : nInner$ **do**

$$\mathbf{f}^{k+1} = (\mu\Phi^*\Phi + \lambda\bar{\mathbf{D}}\bar{\mathbf{D}}^*)^{-1}[\mu\Phi^*(\mathbf{y} - \mathbf{c}^k) + \lambda\bar{\mathbf{D}}(\mathbf{x}^{new} - \mathbf{P}\mathbf{g}^{new} - \mathbf{b}^{new})];$$

$$\mathbf{x}^{k+1} = \text{shrink}(\bar{\mathbf{D}}^*\mathbf{f}^{new} + \mathbf{P}\mathbf{g}^{new} + \mathbf{b}^{new}, 1/\lambda);$$

$$\mathbf{P}\mathbf{g}^{k+1} = \mathbf{P}(\mathbf{x}^{new} - \bar{\mathbf{D}}^*\mathbf{f}^{new} - \mathbf{b}^{new});$$

$$\mathbf{b}^{k+1} = \mathbf{b}^{new} + (\bar{\mathbf{D}}^*\mathbf{f}^{new} + \mathbf{P}\mathbf{g}^{new} - \mathbf{x}^{new});$$

end

$$\mathbf{c}^{k+1} = \mathbf{c}^k + (\Phi\mathbf{f}^{k+1} - \mathbf{y});$$

 Increase k ;

end

References

- [1] R. G. Baraniuk, E. J. Candès, R. Nowak, and M. Vetterli, “Sensing, sampling, and compression,” *IEEE Signal Process. Mag.*, vol. 25, Mar., 2008.
- [2] R. G. Baraniuk, E. J. Candès, M. Elad, and Y. Ma, “Applications of sparse representation and compressed sensing,” *Proc. IEEE*, vol. 98, June, 2010.
- [3] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Constr. Approx.*, vol. 28, pp. 253–263, 2008.
- [4] A. M. Bruckstein, D. L. Donoho, and M. Elad, “From sparse solutions of systems of equations to sparse modeling of signals and images,” *SIAM Rev.*, vol. 51, pp. 34–81, 2009.
- [5] T. Cai, G. Xu, and J. Zhang, “On recovery of sparse signals via ℓ_1 minimization,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 3388–3397, July, 2009.
- [6] E. J. Candès and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, Dec., 2005.

- [7] E. J. Candès, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2006.
- [8] E. J. Candès and T. Tao, “Near optimal signal recovery from random projections: Universal encoding strategies?,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, Dec., 2006.
- [9] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 489–509, Feb., 2006.
- [10] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, “Compressed sensing with coherent and redundant dictionaries,” *Appl. Computat. Harmon. Anal.*, vol. 31, pp. 59–73, 2011.
- [11] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Rev.*, vol. 43, pp. 129–159, 2001.
- [12] O. Christensen, *An Introduction to Frames and Riesz Bases*. Boston, MA: Birkhäuser, 2003, pp. 87–121.
- [13] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Apr., 2006.
- [14] D. L. Donoho, M. Elad, and V. N. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 6–18, Jan., 2006.
- [15] M. Elad, J. L. Starck, P. Querre, and D. L. Donoho, “Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA),” *Appl. Computat. Harmon. Anal.*, vol. 19, pp. 340–358, 2005.
- [16] M. Elad, P. Milanfar, and R. Rubinstein, “Analysis versus synthesis in signal priors,” *Inverse Probl.*, vol. 23, pp. 947–968, 2007.
- [17] T. Goldstein and S. Osher, “The split Bregman algorithm for L1-regularized problems,” *SIAM J. Imag. Sci.*, vol. 2, pp. 323–343, 2009.
- [18] D. Han, K. Kornelson, D. Larson, and Eric Weber, *Frames for Undergraduates*. Providence, RI: American Mathematical Society, 2007.
- [19] W. B. Johnson and J. Lindenstrauss, “Extensions of Lipschitz mappings into a Hilbert space,” *Contemp. Math.*, vol. 26, pp. 189–206, 1984.
- [20] F. Krahmer and R. Ward, “New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property,” *SIAM J. Math. Anal.*, vol. 43, pp. 1269–1281, 2011.
- [21] S. Li, “On general frame decompositions,” *Numer. Funct. Anal. Optim.*, vol. 16, pp. 1181–1191, 1995.
- [22] Y. Liu, T. Mi, and S. Li, “Compressed sensing with general frames via optimal-dual-based ℓ_1 -analysis,” preprint, 2011. [Online]. Available: <http://arxiv.org/abs/1111.4345>
- [23] H. Lütkepohl, *Handbook of Matrices*. New York: Wiley, 1996.
- [24] S. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Trans. Signal Process.*, vol. 41, pp. 3397–3415, Dec., 1993.

- [25] H. Rauhut, K. Schnass, and P. Vandergheynst, “Compressed sensing and redundant dictionaries,” *IEEE Trans. Inf. Theory*, vol. 54, pp. 2210–2219, May, 2008.